Andrew Case

NSS DA7

**Executive Summary**

For my capstone project for the Nashville Software School Data Analytics Cohort, I would like to explore tobacco consumption, the side effects caused by ingesting tobacco, the income it brings in to the farming community in America, and the hole it would leave if the industry were to stop operating.

The United States tobacco industry brings in $12.35 billion a year. It is my belief that this is one of the most corrupt industries in our nation, however without it, our farming production would suffer at alarming rates.

**Motivation**

Since I left college, I have worked with people living in poverty. Many of these individuals struggle with substance abuse. Often these men and women do not receive a stable form of income. Despite all these factors, these men and women still find ways to obtain tobacco products. Addiction is a powerful thing that our country seemingly overlooks due to the revenue tobacco provides.

**Data Question**

 What are the outcomes of tobacco use for each separate item (i.e. cigarettes, pipes, dip, etc.), how much is made from said products, and what would the economic impact be if these items were no longer in circulation?

**Minimum Viable Product (MVP)**

I would like to use Python, PowerBI, and Microsoft Excel to explore the information. I will have no more than three dashboards worth of charts. My target audience is the *Tennessee Tobacco Program, Tennessee Association of Alcohol, Drugs, and other Addiction Services, the CDC, and more.*

**Schedule (through )**

1. Get the Data (11/15/2022)
2. Clean & Explore the Data (11/28/2022)
3. Create Presentation of your Analysis (12/13/2022)

* Should be a presentation, but could include a Jupyter Notebook or dashboard in Excel, Tableau, or PowerBI

1. Internal demos (12/20/2022)
2. Demo Day!! (1/5/2022)

**Data Sources**

[Smoking - Our World in Data](https://ourworldindata.org/smoking)

[Smoking: How large of a global problem is it? And how can we make progress against it? - Our World in Data](https://ourworldindata.org/smoking-big-problem-in-brief)

[Current Cigarette Smoking Among Adults in the United States | CDC](https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm)

**Known Issues and Challenges**

*Explain any anticipated challenges with your project, and your plan for managing them. Be sure to include:*

* *If you need to request data or an api key*
* *Based on your data sources, known data cleaning steps*

The data that I am looking for has several different viewpoints that have been explored. There is a lot of gray area in this topic. For example, several of the men I have worked with recovering from alcohol and drugs would have struggled mightily if they were not allowed to use tobacco. The side effects of tobacco are rarely positive but the economical impact is something that our country might not be able to replace.

Another issue is biases. Information on this topic is rarely unfiltered by one lens or another. My goal is to look across the board to discover what information is simply the facts and do my best to display that data.

Finally, I will be pulling several charts from already made graphs to create and combine the data that I need. This will require web scraping but in a different way than we have explored in class. I look forward to going back to my dear friend Google for some guidance.